

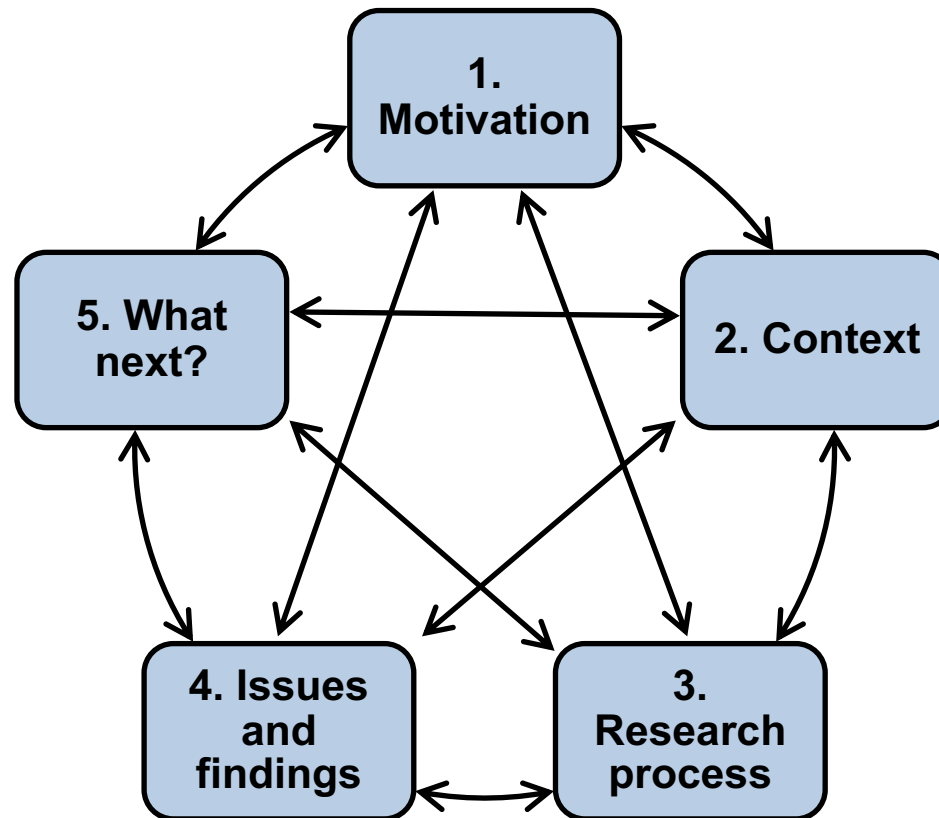
Attributing development impact: lessons from action research with the Qualitative Impact Protocol (QuIP)

Vienna Evaluation Network, May 7, 2020

James Copestake, University of Bath

How do we know whether public interventions are achieving their intended purpose? Impact assessment is a necessary condition for evidence based action but evaluators struggle to do it effectively, particularly in complex and fast changing contexts. This seminar will explore this tension in relation to international development practice by reflecting on action research designing and using a qualitative impact protocol - the 'QuIP' – which aims to be flexible, credible and cost-effective. Drawing on examples of its use from Mexico to India via Mozambique and the UK the presenter will explore ways of addressing confirmation bias, cherry picking, qualitative data analysis and re-docking problems. The seminar draws on “Attributing Development Impact”, available as a free e-book at bit.ly/QuIP-OA.

Outline



1. Motivation – the challenge addressed

Even with effective change monitoring, how can social investors credibly assess their contribution to observed outcomes, particularly in complex contexts?

i.e. **with less**

Enabling intended beneficiaries to voice their views in an **ethical** way

Extractive data collection

Reflecting the **diversity** of their experiences

Over-simplification

Making **cost-effective** investments in finding out more

Obsessive measurement disorder

Additional research questions

How do social investors construct and adapt causal models of change in complex contexts?

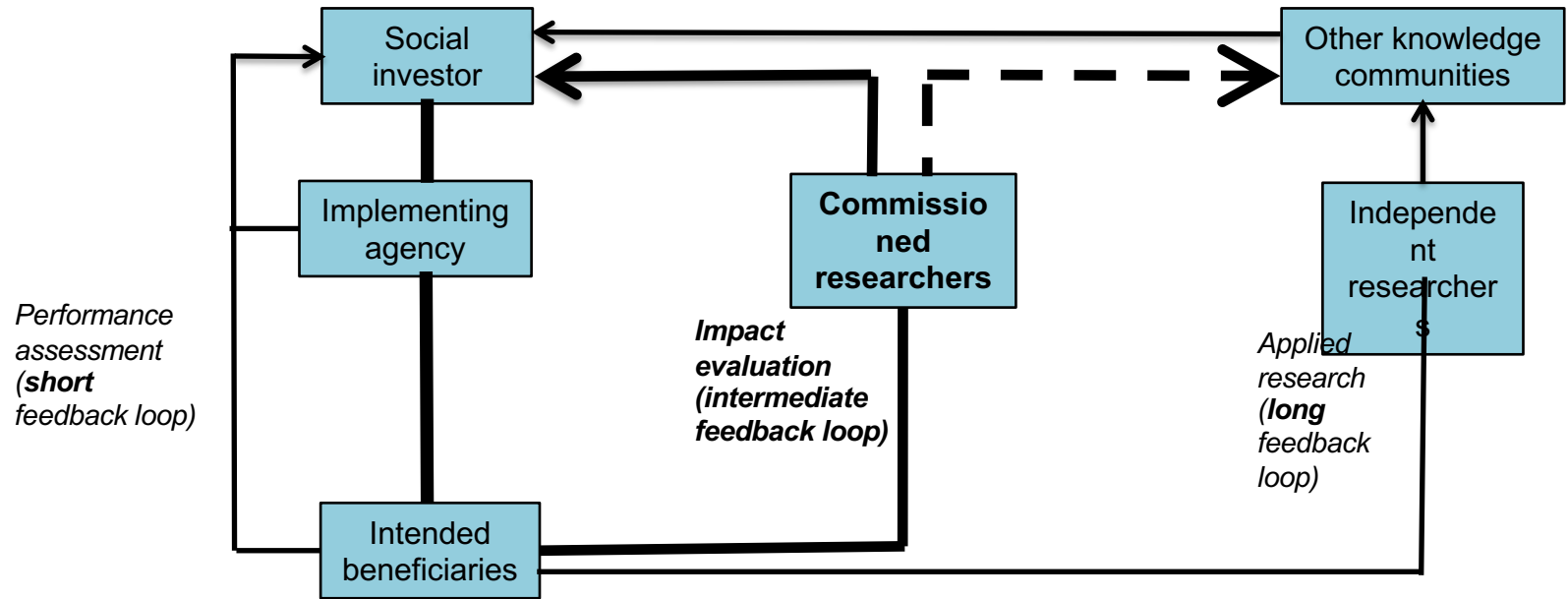
What role can formal evaluation play in evidence-based *political deliberation* over these models?

How is impact evaluation delegitimized?

	Validity or credibility	Usefulness
Change/results monitoring	Key variables can't be measured (immeasurable)	The cost of measuring key variables is too high (imprudent)
Theory based causal claims	Insufficient contextual data (incomplete)	Misdirected, untimely, insufficiently precise (irrelevant)

Adapted from Molecke & Pinkse (2017) Accountability for social impact: a bricolage perspective on impact measurement in social enterprises. *Journal of Business Venturing*.

2. Context



Project specific theories of change –
how (well) will/did this particular
intervention perform?

Mid-range theory / generalisability
- how might similar interventions
perform in other contexts?

General theory
- how does this fit in with wider
understanding?

Mission drift?



*Obsessive measurement
disorder?*

Multiple and contested impact evaluation options

Rely on...	Issues...
Operational data and performance management	Mostly change monitoring, not impact Cognitive biases, vested interests & mission drift Weak external credibility
Variance based impact assessment (observational or experimental, including RCTs)	Costly and indivisible Can be extractive (survey slavery) Narrowly framed (spurious precision, OMD) What works (on average) but not how
Qualitative social research	Akerlof's lemon problem (lack of transparency) Credibility and cost-effectiveness trade-offs
Participatory learning and action	Limited generalisability Limited credibility to non-participants

What a QuIP does not provide

<i>Does not provide...</i>	<i>Responses</i>
Estimates of the <u>magnitude</u> of average treatment effects	Use as one input into microsimulation Run alongside a quantitative impact evaluation
Statistically representative frequency counts	Reveals scope and range of responses Combine with Bayesian updating Use alongside quantitative surveys.
Objective 'facts'	Perceptions matter! Incorporate with other approaches
Recommendations for action	Combine with process evaluation Address in follow-up stakeholder engagement.



3. The research process

Design and pilot testing (2012-15)

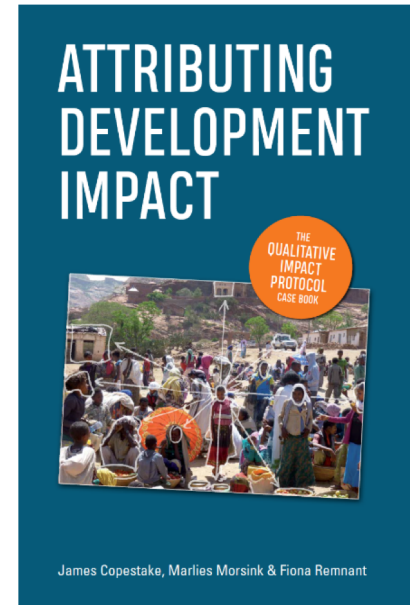
Grant funded collaborative action research to design and test a qualitative approach to impact assessment – tested in Malawi and Ethiopia.

Commercial testing (2016-)

Set up Bath SDR Ltd as a social enterprise to deliver QuIPs in a wider range of contexts. 40+ commissioned QuIP evaluations in 20 countries.

Continued action research (2016-)

Ongoing learning and publication on how to do impact evaluation better, including “*Attributing Development Impact: The QuIP case book*” (2019) - bit.ly/QuIP-OA



BSDR QuIP studies 2016-2019

Activities

Child nutrition
Climate change adaptation
Community mobilisation
Early famine response
Factory working conditions
Housing improvement

Medical & midwife training
Microfinance
Rural livelihoods
Value chain improvement
Sexual & reproductive health rights
Organisational development

Countries

Bolivia	Malawi
Burkina Faso	Mexico
Cameroon	Mozambique
Ethiopia	Nepal
Ghana	Pakistan
Kenya	Sierra Leone
India	Tanzania
Indonesia	Tajikistan
	Uganda
	UK
	Zambia

Commissioners

Acumen	Rutgers Int'l
Bristol City Council	Itad
C&A Foundation	Concern Worldwide
Diageo	Aga Khan Foundation
Self Help Africa	MannionDaniels
Habitat for Humanity	Send a Cow
Oxfam	Oxford Policy Mgt
Save the Children	Power to Change
Seed Global Health	Opportunity Int'l
Tearfund	AgDevCo
Tree Aid	

An overview of the QuIP

If you haven't done so already then please click
[here](#)
and listen to the short video

Features of the QuIP

Self-reported attribution (with latent counterfactuals)
Not statistically inferred attribution based on exposure variation.

Eclectic (contribution analysis, goal-free evaluation, outcome harvesting, most significant change, process tracing, realist evaluation...)

Dual purpose: Exploratory (theory building) and Confirmatory (theory testing).

Bayesian in spirit – augmenting prior theory incrementally; never starting with zero understanding.

More features of QuIP

Designed for **complex contexts** (including presence of multiple causal drivers, measurement problems, rapid change and possible unintended consequences).

Focuses on making **qualitative data analysis** structured, transparent, accessible and useable

Uses **data visualisation – causal maps** - to make reported outcomes more digestible

Good enough – balancing credibility and cost-effectiveness; not aiming for absolute or universal truth.

Combining QuIP with other methods

Supplementing operational data and observation, and process monitoring.

Substituting for other impact evaluation approaches.

Integrating QuIP with other methods (see below)

Before - For exploratory analysis, scoping and identifying variables for quantitative studies.

After - To confirm or deepen understanding of impact claims based on quantitative studies

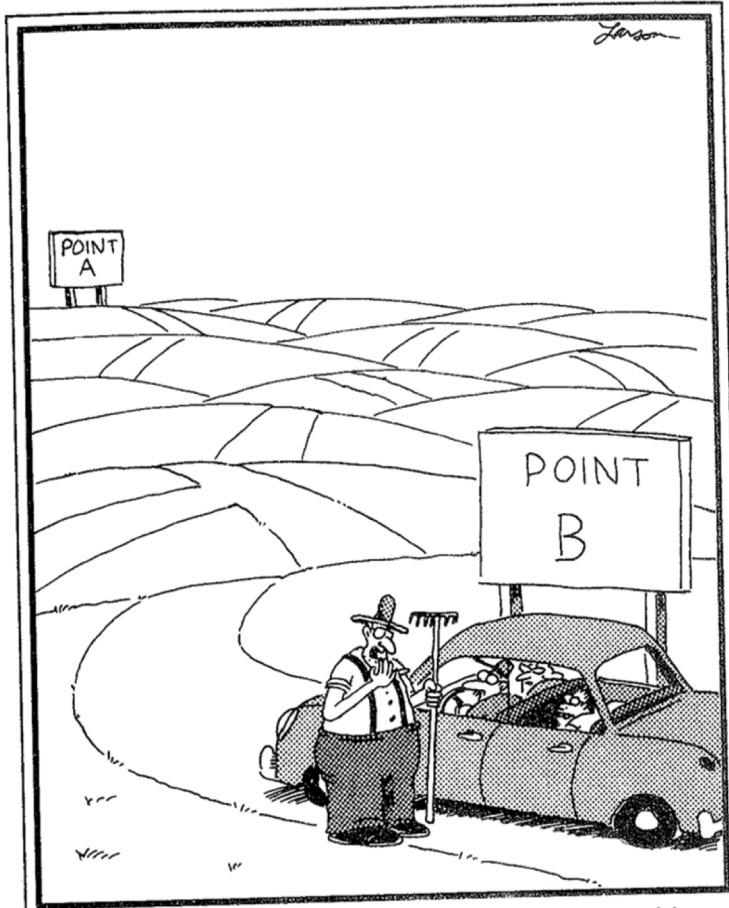
Alongside – To identify mechanisms to explain correlations, instead of relying on theory or speculation.

As an **input** into quantitative analysis (e.g. to inform modelling and simulation).

4. Key methodological issues

Issue	Problem	Solutions
Self-reported attribution	Confirmation bias	Double blindfolding
From field data to synthesis	Opaque data analysis	Thematic (inductive) and attribution (deductive) coding. Dashboard visualisation and cross-referencing numerical & text data.
Robust generalization	Cherry picking of cases and sources	Transparent case selection informed by large 'n' data on context, outcomes (ideally) and prior theory of change.
Effective data use	The re-docking problem	Clarity with the commissioner at the planning stage over scope for deliberation

Blindfolding



"Well, lemme think. ... You've stumped me, son. Most folks only wanna know how to go the other way."

Why do it?

To reduce pro-project, framing and confirmation biases

To give equal weight to all possible drivers of change

Is it ethical?

Informed consent

Time-bound

Greater good (should be proportionate)

How far to go?

A design choice – never zero nor 100%

Only one way of reducing bias

Can be temporary

Can be combined with other sources

Exploratory analysis can also be blindfolded.



Thematic coding and visualization

FILTER

11. Causal chain visual

Source (49)

Target (22)

Attribution Summary (8)

- (All)
- Interesting, no change (9)
- Negative Explicit (2)
- Negative Implicit (4)
- Negative Other (6)
- Positive Explicit (1)
- Positive Implicit (3)
- Positive Other (5)

Gender Code (2)

- (All)
- F
- M

Location Code 1 (2)

- (All)
- E
- H

Location Code 2 (3)

- (All)
- D
- S
- U

Wealth Rank (3)

- (All)
- C
- G
- P

QuIP

11. Causal Chain Visualisation - Respondent Domain Count

Driver of Change

(All)

N1

N2

N3

N5

P1

P10

Outcome 1

(All)

ON1

ON2

ON3

ON4

ON6

ON7

ON9

Outcome 2

(All)

ON1

ON3

ON4

ON5

ON7

ON8

Outcome 3

(All)

ON3

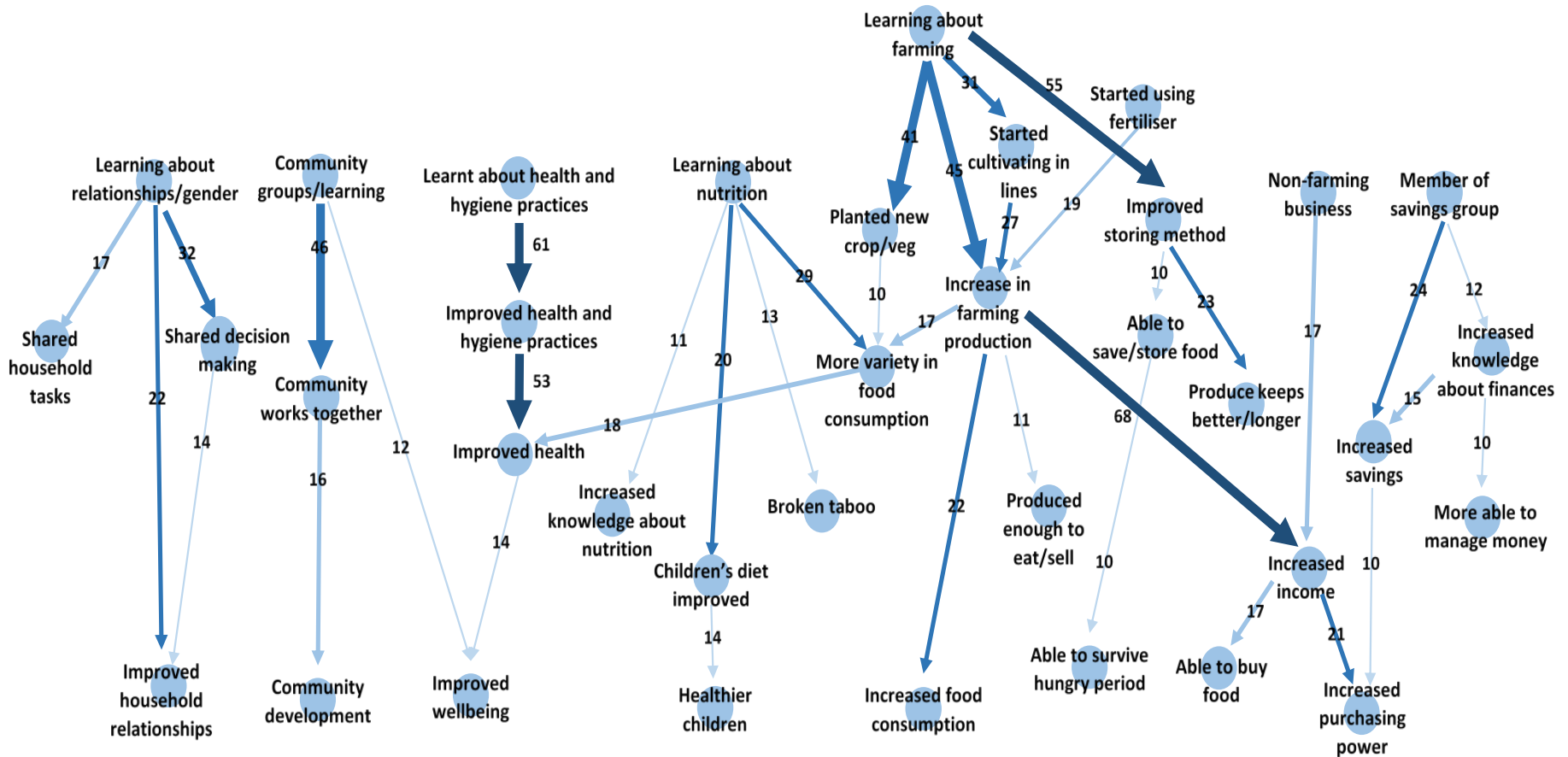
Causal Chain Visualisation

```

graph LR
    A((Increase in food prices)) --> B((Decreased food))
    A --> C((Less money to spend))
    D((Drought)) --> B
    D --> E((More time to find fodder))
    D --> F((No or decreased income))
    G((Delayed PSNP payment)) --> F
    H((End of CFW/ NGO support)) --> F
    I((Livestock no value/ died)) --> F
    B --> C
    B --> J((Sold livestock))
    E --> J
    F --> K((Decreased school attendance))
    F --> L((Disposal of assets))
    J --> M((Worse health children))
    K --> M
    
```

Respondent ID	Question ID	Question Answer
DHFC-2	C1	My husband used to work as daily labor on watershed management to earn money, however, now that activity has ended and we could no more earn money from daily laboring which negatively affect my household income during six months. the livestock has decreased due to drought during the last six months which negatively affect the amount of food and income we make out of them.
	D1	There have been changes in the food items over the last six-months. This is related to both increase in cost of food items and loss of income sources. Purchasing food items are expensive and are very bad changes
	D5	There have been changes in the food items over the last six-months. This is related to both increase in cost of food items and loss of income sources. Purchasing food items are expensive and are very bad changes
DHFC-7	E1	As I have told you, last year drought was bad but different interventions from government organization and NGOs were supporting us. Now there are no more interventions while we are in a recovery period. In addition, I have no income source any more.
	C1	For the last six-months, I have been endeavoring to get food as I spend more money. And this is caused by the increase in price of food items and loss of income source. Happily, I do spend less on accessing clean water. Ultimately, spending more money on
	D1	Last time our income was better than this six months because last time we got different kind of help from different NGOs but now there is no help and we don't have any other income from other source except livestock.
DHFG-2	C1	Concerning the food, the variety of food that we had before is better than during the last six months because we were given different kind of food from different organization as there was drought but now there is no that much food that we get from support. Th hard drought and we were given meat by NGOs. Every family was given a kilo of meat per a day but now there is no meat and we also do not consume different variety of food.
	E1	there is a change on what we spend money during the last six months. Now we spend more money for food than anything than before. It is because last year we got different kind of help including food from different NGOs but now we do not have that help. T
DHFP-10	C1	Previously, we got food support form NGOs and we didn't worry for food but in the last six months we don't get support of food.
	E1	Most of the time we spend more money on food during the last six months than ever before. It is because the price of food items is getting expensive while the value of our livestock has declined because of the drought. Mainly in the last two months we were
DHFP-10	D1	However, the amount food we consumed decreased last six months, because of the high price of the food stuffs. As a result, adults minimized their food intake and priority have been given for children.

Citation Count: 10+



Evidence selection – definitions and principles

Saturation – Maximize potential to learn about the fullest possible range of drivers of change affecting the selected population

(Exploratory)

Bayesian updating - Maximize the potential to test prior theory

(Confirmatory)

Heterogeneity and differential impact - Capture the most important sources of variation (see below).

Equating marginal benefits and costs – Invest in more data until the extra evidence does not justify the extra cost.

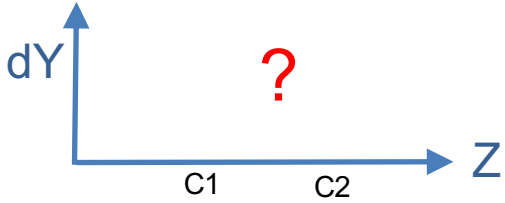
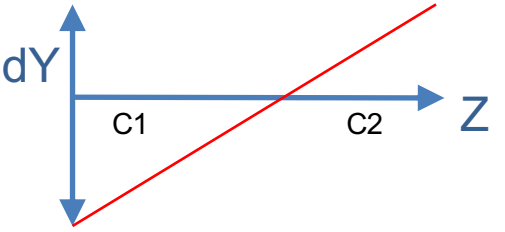
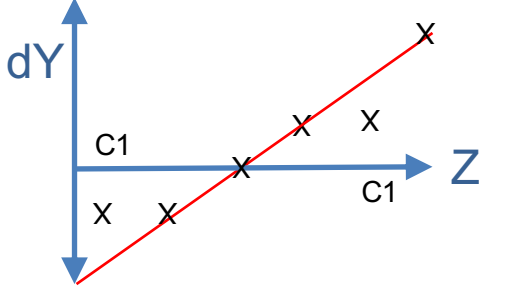
- The principles for doing this are NOT the same as those for estimating the average value of a known variable across a population
- Random selection is NOT necessarily best, indeed is unlikely to be.
- Doing all data at the same time is NOT necessarily optimal.
- There is no scientifically optimal sample size, but transparent and reasoned case selection is important to rebuff criticisms of ‘cherry-picking’.

Evidence selection options (Linked to availability of prior data)

Option	Treatment data (X)?	Outcome data (Y)?	Contextual data (Z)?	Comment
A	No	No	No	Random selection across full population is the only option
B	Yes	No	No	Select randomly from quota samples across categories of treatment or exposure
C	No	Yes	No	Select purposively to include positive and negative deviants
D1	No	No	Yes	Select purposively to reflect important dimensions of variation across the population (e.g. gender, age)
D2	No	No	Yes	Select purposively to include likely positive and negative deviants according to prior theory.
E	No	Yes	Yes	Select purposively to include anomalous cases poorly explained by prior theory linking Z and Y.

Case selection – illustration

(Two cluster source data selection scenarios for a uniform X)

1		No data for dY (outcome) or theory of change linking it to context (Z)	Maximise dispersion of cases across Z, or select randomly
2		Theory of change linking Z to dY, but no data for dY	Choose expected negative (C1) and positive (C2) deviant cases
3		Data for Z and dY permits empirical testing of the theory	Select on anomalous cases or outliers relative to theory: positive (C1) and negative (C2)

The redocking problem

Work back from use

To stimulate internal learning
(and consultation with a wider diversity of stakeholders)

To influence operational decisions (e.g. mid-project reviews and adjustment, scaling up, closing or adapting activities)

To support wider deliberation
over core theories of change

For external reporting (with need for sufficient credibility to match potential push back)

via outputs

A written **report** of findings

An interactive **dashboard**

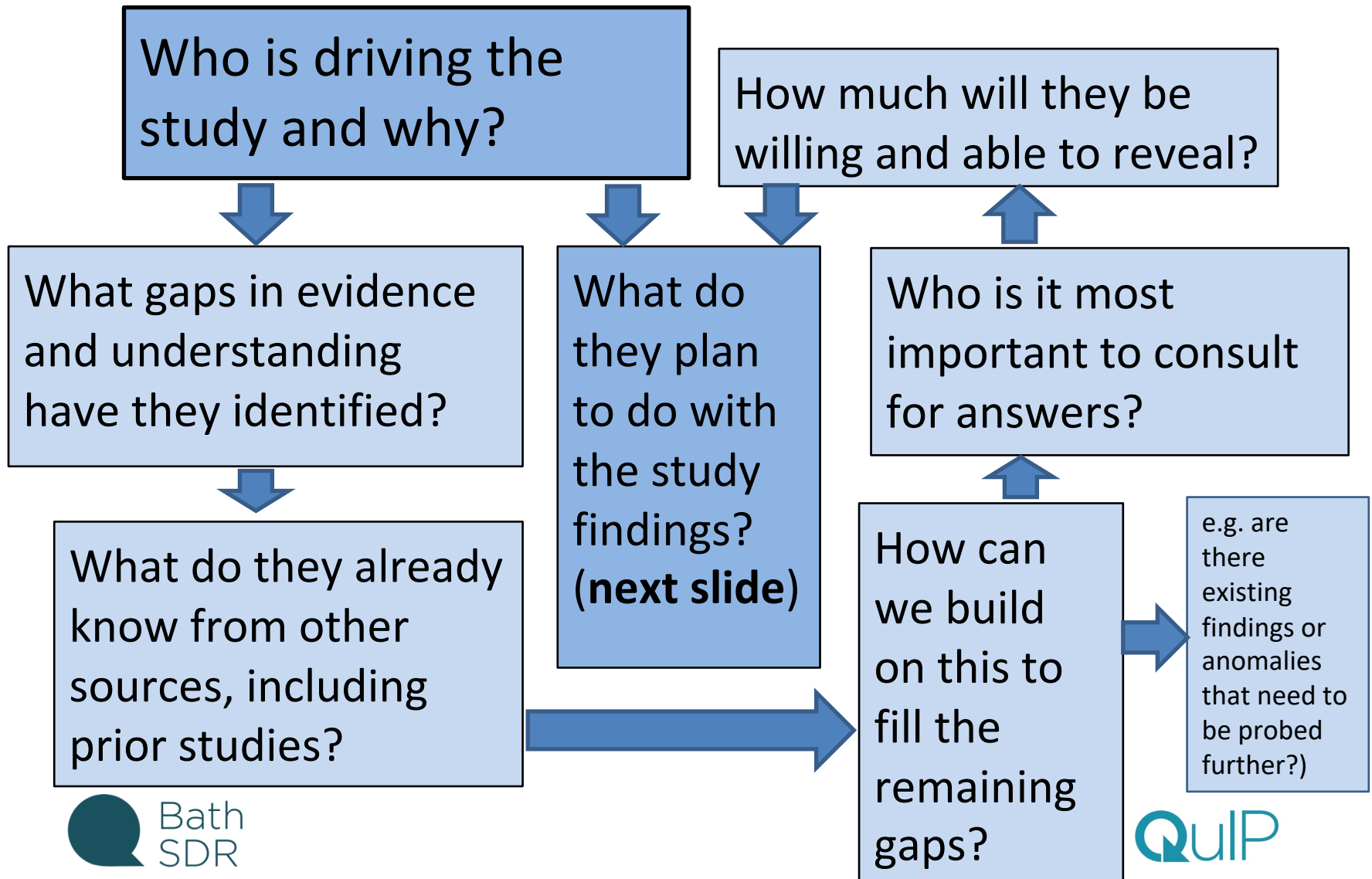
Integration with other sources of evidence

Participatory sense-making **events**

Media based wider dissemination



The key to re-docking – initial framing

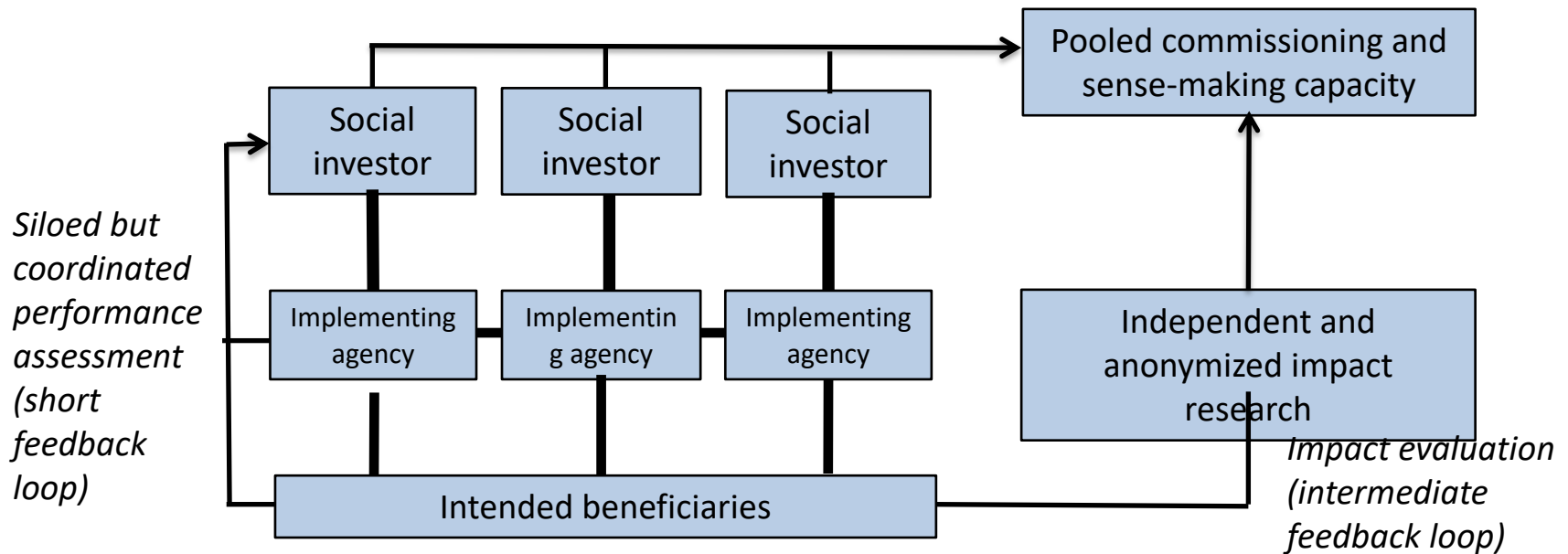


What next? Evaluating complex problems

1. Elaborate the intervention **theory of change**
2. Monitor changes in key outcome indicators and their possible causal drivers across the population (**large 'n'** or wide-but-shallow data).
3. Ask key informants to identify credible causal mechanisms between observed changes (**small 'n'** or narrow-but-deep data).
4. Collect addition specific evidence to adjust these prior causal claims.
5. Iterate between these steps until a threshold of sufficiently credible evidence is achieved (or we run out of time and money).



Pooled commissioning to address the silo problem



Bath SDR: QuIP theory of change

